

# An information theoretic measure for the evaluation of ordinal scale data

W. J. TASTLE

*Ithaca College, Ithaca, New York*

and

M. J. WIERMAN

*Creighton University, Omaha, Nebraska*

This article describes a new measure of dispersion as an indication of consensus and dissention. Building on the generally accepted Shannon entropy, this measure utilizes a probability distribution and the ordered ranking of categories in an ordinal scale distribution to yield a value confined to the unit interval. Unlike other measures that need to be normalized, this measure is always in the interval 0 to 1. The measure is typically applied to the Likert scale to determine degrees of agreement among ordinal-ranked categories when one is dealing with data collection and analysis, although other scales are possible. Using this measure, investigators can easily determine the proximity of ordinal data to consensus (agreement) or dissention. Consensus and dissention are defined relative to the degree of proximity of values constituting a frequency distribution on the ordinal scale measure. The authors identify a set of criteria that a measure must satisfy in order to be an acceptable indicator of consensus and show how the consensus measure satisfies all the criteria.

Ordinal scales of measurement typically consist of ordered category hierarchies, such as *strongly agree, agree, neither agree nor disagree, disagree, and strongly disagree* or *very cold, cold, cool, tepid, warm, hot, and very hot*. The instrument typically used to collect this kind of data is called the *Likert scale*, although there are variations of this scale, such as Likert-like, Likert-type, and ordered response scales. Researchers utilize this kind of instrument to collect data that cannot be ascertained using traditional measures, for the data being collected are feelings, perceptions, sensations, emotions, impressions, sentiments, opinions, passions, or the like. Unfortunately, the application of standard statistics to these data can be improper (Cohen, Manion, and Morrison, 2000; Jamieson, 2004; Pell, 2005). This article looks at the different kinds of scales and presents a new measure for analyzing ordinal scale data.

## The Use of Mathematics in Measures

Stevens (1946) introduced four levels of measurement: nominal, ordinal, interval, and ratio.

Collections of categories used to accumulate data that are without any sense of order are called *nominal* scales. An example of a nominal scale is a listing of continents—for example, 1 = Africa, 2 = Antarctica, 3 = Asia.

*Ordinal* scales are merely *ordered* categories, but the ordering makes comparisons implicitly possible. Testing

the temperature of a cup of tea would permit someone to use the words *cold, tepid, lukewarm, warm, and hot* as their scale of comparative measure. There is no sense of interval scale in this measure, and hence, equations such as  $cold + lukewarm = hot$  or the average of warm and hot is warm and a half are impractical and illogical. Likert scales fall into this category of measures. The number of choices that may be selected as categories of a Likert scale are virtually without limit, although five or seven category scales are the most prominent. The distance between each category, sometimes referred to as the *interval*, is incorrectly assumed to be equal, although the names of the categories chosen do attempt to create some sense of quasiequal intervalness. Nevertheless, there is no empirical evidence to suggest that the interval between *neutral* and *agree* is equal to the interval between *strongly disagree* and *disagree*.

*Interval* scales possess a definite and fixed interval between consecutive values. *Ratio* scales have an absolute zero base, possess an interval, and implicitly possess order. The number line is a ratio scale, and all mathematical operations can be conducted on such a scale.

The dispersion of values about a central value—that is, the mean—permits an assessment of the strength of the collective respondents' perceptions without placing a focus on an arbitrary numerical interval assignment. Thus, a collective set of ordinal scale values that yield a narrow dispersion can logically be viewed as possessing a greater agreement than does one with a wide dispersion. The logic is identical to that of the standard deviation. As will be shown below, the consensus measure informs the investigator of the sense of dispersion, using the commonly understood concept of percentage.

---

Correspondence concerning this article should be addressed to W. J. Tastle, Department of Business Administration, Ithaca College, Ithaca, NY 14850 (e-mail: [tastle@ithaca.edu](mailto:tastle@ithaca.edu)).

The mean requires a fixed interval and a continuous scale, neither of which is available in an ordinal scale. Therefore, we use a measure of dispersion that does not require a fixed interval or continuous scale but does offer a minimum and maximum value. By using a relative measure, such as a percentage, there need not be any agreed-to interval scale; all that needs to be decided is the location of the extreme values associated with the endpoints of the range of categories; in the case of the Likert scale, the extrema are *strongly agree* and *strongly disagree*.

The consensus measure satisfies the considerations above. It imposes an interval scale upon the ordinal data but is only weakly dependent on these intervals, and as such, it is a suitable measure of agreement when applied to ordinal scales. The original motivation for the measure was to resolve a problem dealing with group decision-making dynamics and was thus named the *consensus* measure.

### RULES FOR A MEASURE OF CONSENSUS

We establish a set of rules that must be satisfied before any measure can be considered a viable solution to the ordinal scale agreement/consensus problem.

1. For a given (even) number,  $n$ , of datum points used to collect information on some question of interest that is represented in the form of a Likert or Likert-type scale, if an equal number of data,  $n/2$ , separate themselves into two disjoint groupings, each centered on the *strongly disagree* and *strongly agree* categories (or any other such extrema), the data set is considered to have no agreement/consensus. For a given odd number of datum points, there is at least one datum that is not on an extreme, forcing the consensus to be greater than zero.

2. If all the data points are classified in the same category, regardless of the category, the consensus of the data is considered to be complete at 100%.

3. If the mix of data points is such that exactly  $(n/2) - 1$  points are assigned to any one category, the degree of consensus must be greater than 0, for the balance among the data is no longer equal.

4. As the categories containing data converge (away from the extremes of the scale) to a single category, the consensus must increase, eventually approaching 1. This is representative of the breakdown in a bimodal distribution in which the perceptions of the individuals assigning their views to the categories are less diverse. Consequently, as the views of the survey responders become less in their extreme views and more accommodating of the views of others, the frequency distribution of the values associated with the ordered categories contains representations of more categories, instead of the extreme two. Thus, the frequency distribution converges toward a single, agreed-to category. When all survey responders are in complete agreement, consensus is attained, and that is given a value of 1.

### BRIEF HISTORICAL OVERVIEW

The motivation for this measure came out of a need to identify a means by which ordinal data, in the form of

the Likert scale, could be used to provide a measure of consensus, so that the progress, or lack thereof, of a group of individuals as they sought to attain consensus on some question could be given a meaningful number. The number could then be compared against other numbers taken at previous times during the discussion, to determine whether the group was headed toward an eventual consensus. Thus, the history of its development centers on group decision making, as this brief overview will demonstrate, although the application of this measure to the solution of other kinds of problems has continued to grow.

Herrera, Herrera-Viedma, and Verdegay (1996) examined consensus from the perspective of linguistic labels: *certain*, *extremely likely*, *most likely*, *meaningful chance*, *it may*, *small chance*, *very low chance*, *extremely unlikely*, and *impossible*. After considerable matrix computation, they determined that a consensus measure is determined by degrees: level of preference, level of alternative, and level of relation. A typical group might be challenged in determining how to utilize this measure.

Tcha, Ahammad, and Qiang (2002) wrote an interesting paper analyzing Ph.D. student reflections, using the following equation to assign a value from the unit interval as a measure of consensus. Initially, this appears to be an application of the Shannon entropy equation:

$$\text{Consensus} = 1 - \frac{\sum p_i \times \ln(p_i)}{\ln(1/n)},$$

where  $n$  is the number of categories. However, the measure fails upon closer inspection. One of the attributes of the Shannon entropy is its ability to measure the amount of uncertainty associated with any probability distribution. Hence, given identical probability distributions of  $n$  categories, there are  $n!$  ways in which the categories can be ordered within each distribution. Each separate ordering of the distribution will have the same entropy value associated with each, for it is a principle of entropy that the ordering of the categories within a distribution should not effect the entropy value. Entropy is constant regardless of the order of the categories within the distribution. This is exactly in *opposition* to the requirements essential to a consensus measure, for as the shape of the distribution changes, so must the measure of consensus reflected by that change.

The Shannon entropy, denoted in the literature as either  $S(X)$  or  $H(X)$ , is

$$S(X) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i),$$

where  $X$  is the set of  $n$  categories under investigation and  $p(x_i)$  is the probability of each  $x_i$ . A comprehensive description of this measure and its properties can be found in Klir and Wierman (1997). Since the Shannon entropy already possessed some of the properties desired in the consensus measure, we chose to focus our research on an enhancement of this work.

Szmidt and Kacprzyk (2003) have provided an interesting look at consensus from the perspective of individual

intuitionistic fuzzy preference relations by which a distance from consensus can be determined. Using the unit interval, they defined complete disagreement as equal to 0 and complete agreement as 1. This is superior to the method of merely using a mean and some traditional measure of variance to determine proximity to consensus, for those values vary with the number of participants.

**THE MEASURES OF CONSENSUS AND DISSENTION**

Let us assume that we have a five-attribute Likert scale: *strongly agree* (SA), *agree* (A), *neutral* (N), *disagree* (D), and *strongly disagree* (SD). Let us further assign a numerical scale of SA = 1, A = 2, N = 3, D = 4, and SD = 5. Then  $X = \{1, 2, 3, 4, 5\}$  and  $X_1 = 1, X_2 = 2$ , and so on. The width of  $X, d_X$ , is  $X_{max} - X_{min}$ . In this case,  $d_X = 5 - 1 = 4$ . The mean, or expected value, of  $X$  is given by the usual formula

$$E(X) = \sum_{i=1}^n p_i X_i = \mu_X.$$

Table 1 illustrates 10 different results that might be attained from a population of 12 individuals, using a five-category Likert scale.

The consensus measure is defined (Wierman & Tastle, 2005; Tastle & Wierman, 2005) as

$$Cns(X) = 1 + \sum_{i=1}^n p_i \log_2 \left( 1 - \frac{|X_i - \mu_X|}{d_X} \right),$$

where  $X$  represents the Likert scale,  $X_i$  is a particular Likert attribute,  $p_i$  is the probability or frequency associated with each  $X_i$  (the relative number of individuals who gave that response),  $d_X$  is the width of  $X$ , and  $\mu_X$  is the mean of  $X$ .

Dissent is defined as  $1 - Cns(X)$  and will be discussed in the next section. However, the reader should keep in mind that both consensus and dissention are intimately related and that, given one, the other is easily calculated by  $Cns(X) = 1 - Dnt(X)$ .

The measure Cns of consensus adequately satisfies the rules in the Rules for a Measure of Consensus section, as corroborated by the following *illustrations* (the proof is in

the Range of Consensus and Dissention section and the Appendix).

Refer again to Table 1. The first column simply numbers the 10 different results. The next five columns show how many of the 12 individuals chose each response in each of the 10 cases. The sixth column provides the mean of each result, under the assumption that SA = 1, A = 2, N = 3, D = 4, and SD = 5. The mean is the average for the category values, so a mean of 2.5 represents a point halfway between *agree* and *neutral*, Cns and Cns% are the consensus values in decimal and rounded percent, and Dnt and Dnt% are the dissention values in decimal and rounded percent.

Row 1 shows a maximum amount of dissent in agreement, since  $n/2$  values are reflected in the extreme attributes. As a point of interest, had the  $n/2$  values been associated with *agree* and *disagree*, the consensus would have been 0.5850 since these attributes are closer to each other. Rows 2–9 show a convergence of dispersion moving toward *agree*. An examination of the means shows a modest fluctuation of the values but, in general, a movement of value from 3 = *neutral* to 2 = *agree*. The consensus shows continuous movement toward 1; it is arguably easier to associate the consensus as a percentage, in order to easily visualize the movement toward an agreement.

Conversely, one can monitor the dissent from complete (row 1) to absence (row 10). Finally, row 10 shows the attribute values firmly in one category. The mean is expectedly at 2, consensus is complete at 100%, and dissent does not exist. Of particular interest are rows 7 and 8, in which the mean is equal but the consensus differs. This illustrates the use of the mean as questionable, by itself, in analyzing ordinal data.

**THE RANGE OF CONSENSUS AND DISSENTION**

The information measure takes a histogram, or probability distribution, over a discrete set of choices with known weights and produces a single value that ranges from 0 (*complete disagreement*) to 1 (*complete agreement*). The following establishes definitions and examines some of the mathematical properties of the consensus measure.

**Table 1**  
Illustration of 10 Sets of Values Ranging From the Most Extreme (Row 1) to the Most Concentrated (Row 10)

	SA	A	N	D	SD	Mean	Cns	Cns%	Dnt	Dnt%
1	6	0	0	0	6	3.0	0	0	1	100
2	6	0	0	1	5	2.9167	0.0492	5	0.9508	95
3	6	0	0	2	4	2.8333	0.0996	10	0.9004	90
4	5	1	0	2	4	2.9167	0.1468	15	0.8532	85
5	5	1	2	4	0	2.4167	0.4437	44	0.5563	56
6	1	5	2	4	0	2.7500	0.6103	61	0.3897	39
7	1	5	4	2	0	2.5833	0.6866	69	0.3134	31
8	0	6	5	1	0	2.5833	0.7676	77	0.2324	23
9	0	9	3	0	0	2.2500	0.8553	86	0.1447	14
10	0	12	0	0	0	2.0	1	100	0	0

Note—SA, strongly agree; A, agree; N, neutral; D, disagree; SD, strongly disagree; Cns, consensus; Cns%, consensus as a percentage; Dnt, dissent; Dnt%, dissent as a percentage.

**Definition 1.** Let  $X$  be a discrete random variable of size  $n$  with probability distribution  $p(X)$ . As usual,  $\mu_X$  is the mean of  $X$ . Let  $d_X = X_{\max} - X_{\min}$  be the width of  $X$  and set  $d_i = |X_i - \mu_X|$  as the absolute deviation of  $X$  from the mean. The consensus,  $\text{Cns}(X)$ , is then defined to be

$$\begin{aligned} \text{Cns}(X) &= 1 + \sum_{i=1}^n p_i \log_2 \left( \frac{d_X - d_i}{d_X} \right) \\ &= 1 + \sum_{i=1}^n p_i \log_2 \left( 1 - \frac{|X_i - \mu_X|}{d_X} \right). \end{aligned} \tag{1}$$

If there is no chance of confusion, we will drop the subscripts and write

$$\text{Cns}(X) = 1 + \sum_{i=1}^n p_i \log_2 \left( 1 - \frac{|X_i - \mu|}{d} \right). \tag{2}$$

The mirror image of consensus is *dissention* and has the following form:

$$\begin{aligned} \text{Dnt}(X) &= - \sum_{i=1}^n p_i \log_2 \left( \frac{d_X - d_i}{d_X} \right) \\ &= - \sum_{i=1}^n p_i \log_2 \left( 1 - \frac{|X_i - \mu_X|}{d_X} \right). \end{aligned} \tag{3}$$

Again, if there is no chance of confusion, we will drop the subscripts and write

$$\text{Dnt}(X) = - \sum_{i=1}^n p_i \log_2 \left( 1 - \frac{|X_i - \mu|}{d} \right). \tag{4}$$

Dissention is one minus the consensus,  $\text{Dnt}(X) = 1 - \text{Cns}(X)$ ; the main mathematical result of the remaining portion of the section will justify this definition. Dissention has the usual form of a measure of information, and this article will use the dissention in the proofs that follow.

**Linear Transform and Dissention**

**Lemma 2.** Let  $X$  be a discrete random variable of size  $n$  with probability distribution  $p(x)$ . Let  $d_X = X_{\max} - X_{\min}$  be the width of  $X$ . We define

$$Y_i = \frac{X_i - X_{\min}}{d_X}, \tag{5}$$

and it follows that  $\text{Dnt}(Y) = \text{Dnt}(X)$ .

*Proof.*

$$\text{Dnt}(Y) = - \sum_{i=1}^n p_i \log_2 \left( 1 - \frac{|Y_i - \mu_Y|}{d_Y} \right), \tag{6}$$

but since Equation 5 is a linear transform that preserves order, we know that

$$\begin{aligned} d_Y &= Y_{\max} - Y_{\min} \\ &= \frac{X_{\max} - X_{\min}}{X_{\max} - X_{\min}} - \frac{X_{\min} - X_{\min}}{X_{\max} - X_{\min}} \\ &= 1 \end{aligned} \tag{7}$$

and

$$\mu_Y = \frac{\mu_X - X_{\min}}{d_X}. \tag{8}$$

If we substitute these two expressions, as well as the definition of  $Y_i$  into  $\text{Dnt}(Y)$ , we get

$$\begin{aligned} \text{Dnt}(Y) &= - \sum_{i=1}^n p_i \log_2 \left( 1 - \frac{\left| \frac{X_i - X_{\min}}{d_X} - \frac{\mu_X - X_{\min}}{d_X} \right|}{1} \right) \\ &= - \sum_{i=1}^n p_i \log_2 \left( 1 - \frac{|X_i - \mu_X|}{d_X} \right) \\ &= \text{Dnt}(X), \end{aligned} \tag{9}$$

since  $d_X$  is positive.

From this lemma, we see that as long as  $n \geq 2$ , we can assume that the values of the random variable  $Y$  are in the unit interval and that  $Y_{\min} = 0$  and  $Y_{\max} = 1$ . For this transformed random variable,  $d_Y = 1$ , and thus, we have the following corollary.

**Corollary 3.** If  $X$  is a discrete random variable as given in the transform lemma and  $Y$  is given by Equation 5, then

$$\begin{aligned} \text{Dnt}(X) &= \text{Dnt}(Y) \\ &= - \sum_{i=1}^n p_i \log_2 (1 - |Y_i - \mu_Y|). \end{aligned} \tag{10}$$

In all of the following, we will assume that any random variable  $X$  has been transformed using Equation 5, so that the values of  $Y$  span the unit interval, and that the dissention is, therefore, given by Equation 10.

**Bounds of the Dissention Measure**

We show that the dissention is bounded below by zero and above by one. This immediately implies that its mirror, the consent, given in Equation 1, is also bounded below by zero and above by one. Trivially, the dissention measure is positive, since each individual term in the summation given by Equation 10 is positive.

**Lemma 4.** If  $0 < \mu < 1$  and  $-\mu < \varepsilon < 1 - \mu$ , then  $0 < \mu - \varepsilon, \mu + \varepsilon < 1$ , and

$$\mu^{\mu+\varepsilon} (1 - \mu)^{1-\mu-\varepsilon} \leq 1 - |\varepsilon|. \tag{11}$$

*Proof.* See the Appendix for the proof of Lemma 4.

**Corollary 5.** If  $0 < \mu < 1$  and  $-\mu < \varepsilon < 1 - \mu$ , then

$$\begin{aligned} & - \log_2 (1 - |\varepsilon|) + (\mu + \varepsilon) \log_2 \mu \\ & + (1 - \mu - \varepsilon) \log_2 (1 - \mu) \leq 0. \end{aligned} \tag{12}$$

*Proof.* Take the log of both sides of Equation 12 and perform some algebraic manipulation.

**Theorem 6.** If  $n = 2$  is the size of the random variable  $Y$ , then  $\text{Dnt}(Y) \leq 1$ .

*Proof.* See the Appendix for the proof of Theorem 6.

Now that we have a base case, we shall prove by induction that the dissention is always less than or equal to one for the transformed random variable  $Y$ .

**Theorem 7.** *If  $Y$  is a discrete random variable of size  $n \geq 2$ , with  $Y_{\min} = 0$  and  $Y_{\max} = 1$ , then  $\text{Dnt}(Y) \leq 1$ .*

*Proof.* See the Appendix for the proof of Theorem 7.

### RELATING DISSENTION TO OTHER STATISTICS

The consensus and dissent measure have been defined as expected values. The dissent is the expected value of the log of one minus the distance of  $Y_k$  from the mean,

$$\text{Dnt}(Y) = E\left[-\log_2\left(1 - |Y_i - \mu_Y|\right)\right].$$

It is known that the expected value of the squared distance

$$E\left[(Y_i - a)^2\right]$$

is minimized if  $a = \mu_Y$ , the mean, and that the expected value of the absolute distance

$$E\left[|Y_i - a|\right]$$

is minimized if  $a = m_Y$ , the median. The squared distance is used because it is easier to deal with analytically. Empirical results have shown that neither the mean nor the median minimize the expression

$$\text{Dnt}(Y) = E\left[-\log_2\left(1 - |Y_i - a|\right)\right],$$

viewed as a function of  $a$ . However, these same empirical results have shown that the mean of  $Y$  generally gives a lower value than does the median to the dissent.

### CONCLUSION

The consensus measure offers statisticians, psychologists, organizational behaviorists, social scientists, and researchers in a plethora of other disciplines a tool by which ordinal scales can be compared and contrasted. This new measure is equally applicable to interval and ratio scales; research into its usage is currently underway in various areas of interest. An interesting development is the softening of the requirement that the category values be integers. The measure works equally well under situations of ambiguity when the investigator cannot commit to one single action: Perhaps the instance being measured is mostly of

one category but part of another. Using real numbers in the category values permits the measure to be applied to another class of problems, with the single provision that the fractional parts sum to one.

The new measure of consensus serves as a measure of agreement, as well as the inverse of dissent. Using a measure devised from the discipline of information theory, it is easy to develop an intuition for dispersion of data, be it ordinal, interval, or ratio, in a manner that is impossible to ascertain with conventional statistics. Equally possible is the ability to compare measures of a data set over time in order to determine whether the investigation being undertaken at hand, be it group dynamics, some activity in nature, or a laboratory experiment, is approaching consensus or agreement or dissent when the data being measured are qualitative in nature and the scale is ordinal, interval, or ratio.

### REFERENCES

- COHEN, L., MANION, L., & MORRISON, K. (2000). *Research methods in education* (5th ed.). London: Routledge Falmer.
- HERRERA, F., HERRERA-VIEDMA, E., & VERDEGAY, J. L. (1996). A model of consensus in group decision making under linguistic assessments. *Fuzzy Sets & Systems*, **78**, 73-87.
- JAMIESON, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, **38**, 1217-1218.
- KLIR, G. J., & WIERMAN, M. J. (1997). *Lecture notes in fuzzy mathematics and computer science: Uncertainty-based information elements of generalized information theory*. Omaha: Creighton University, Center for Research in Fuzzy Mathematics and Computer Science.
- PELL, G. (2005). Use and misuse of Likert scales. *Medical Education*, **39**, 970.
- STEVENS, S. S. (1946). On the theory of scales of measurement. *Science*, **103**, 677-680.
- SZMIDT, E., & KACPRZYK, J. (2003). A consensus-reaching process under intuitionistic fuzzy preference relations. *International Journal of Intelligent Systems*, **18**, 837-852.
- TASTLE, W. J., & WIERMAN, M. J. (2005). Consensus and dissent: A new measure of agreement. In *Annual Meeting of the North American Fuzzy Information Processing Society, 2005 (NAFIPS 2005)* (pp. 385-388). Los Alamitos, CA: IEEE Press.
- TCHA, M., AHAMMAD, H., & QIANG, Y. (2002). *What future Australian professors in economics and business think: Results from twin surveys of PhD students*. Available at [www.econs.ecel.uwa.edu.au/](http://www.econs.ecel.uwa.edu.au/).
- WIERMAN, M. J., & TASTLE, W. J. (2005). Consensus and dissent: Theory and properties. In *Annual Meeting of the North American Fuzzy Information Processing Society, 2005 (NAFIPS 2005)* (pp. 75-79). Los Alamitos, CA: IEEE Press.

**APPENDIX**  
**Proofs**

**Lemma 4.** *If  $0 < \mu < 1$  and  $-\mu < \varepsilon < 1 - \mu$ , then  $0 < \mu - \varepsilon, \mu + \varepsilon < 1$ , and*

$$\mu^{\mu+\varepsilon} (1 - \mu)^{1-\mu-\varepsilon} \leq 1 - |\varepsilon|. \tag{A1}$$

*Proof.* If  $\varepsilon = 0$ , then  $\mu^\mu(1 - \mu)^{1-\mu}$  is certainly less than 1. If we fix  $\mu$  and differentiate  $f(\varepsilon) = \mu^{\mu+\varepsilon}(1 - \mu)^{1-\mu-\varepsilon}$  with respect to  $\varepsilon$  twice, we get

$$\frac{df}{d\varepsilon} = \mu^{\mu+\varepsilon} (1 - \mu)^{1-\mu-\varepsilon} [\log_2(\mu) - \log_2(1 - \mu)] \tag{A2}$$

and

$$\frac{d^2f}{d\varepsilon^2} = \mu^{\mu+\varepsilon} (1 - \mu)^{1-\mu-\varepsilon} [\log_2(\mu) - \log_2(1 - \mu)]^2. \tag{A3}$$

If  $\mu = 1 - \mu = 1/2$ , then both the first and the second derivatives are zero everywhere, and  $f(\varepsilon) = 1/2$ , a horizontal line. The situation is illustrated in Figure A1. Otherwise, the second derivative is always positive, which tells us that the curve  $f$  is convex. There are two solutions to  $\mu^{\mu+\varepsilon}(1 - \mu)^{1-\mu-\varepsilon} = 1 - |\varepsilon|$  at  $\varepsilon = -\mu$  and  $\varepsilon = 1 - \mu$ . Due to convexity, the curve of  $f$  must be lower than  $1 - |\varepsilon|$  over the entire interval  $(-\mu, 1 - \mu)$ , since it is lower at zero, an interior point. The situation is illustrated in Figure A2. When  $\mu \neq 0.5$ , then  $\mu^{\mu+\varepsilon}(1 - \mu)^{1-\mu-\varepsilon}$  is convex and below  $1 - |\varepsilon|$  over the interval  $-\mu < \varepsilon < 1 - \mu$ .

**Theorem 6.** *If  $n = 2$  is the size of the random variable  $Y$ , then  $\text{Dnt}(Y) \leq 1$ .*

*Proof.* If  $n = 2$ , then  $Y_{\min} = 0, Y_{\max} = 1, d = Y_{\max} - Y_{\min} = 1$ , and  $\mu = p_1 Y_{\min} + p_2 Y_{\max} = p_2$ . We can assume, without loss of generality, that  $0 < p_1 < 1$  for  $i = 1, 2$ . Since  $\mu = p_2$ , we know that  $|1 - \mu| = 1 - p_2 = p_1$  and  $|0 - \mu| = p_2$ , so that

$$\begin{aligned} \text{Dnt}(Y) &= - \sum_{i=1}^2 p_i \log_2(1 - |Y_i - \mu_Y|) \\ &= - \sum_{i=1}^2 p_i \log_2(1 - |Y_i - p_2|) \\ &= - p_1 \log_2(1 - |0 - p_2|) - p_2 \log_2(1 - |1 - p_2|) \\ &= - p_1 \log_2(p_1) - p_2 \log_2(p_2). \end{aligned} \tag{A4}$$

This last equation is exactly of the form of the Shannon entropy for  $n = 2$  probabilities. It is well known that Shannon entropy for  $n = 2$  reaches its maximum of  $\log_2 2 = 1$  when  $p_1 = p_2 = 1/2$ .

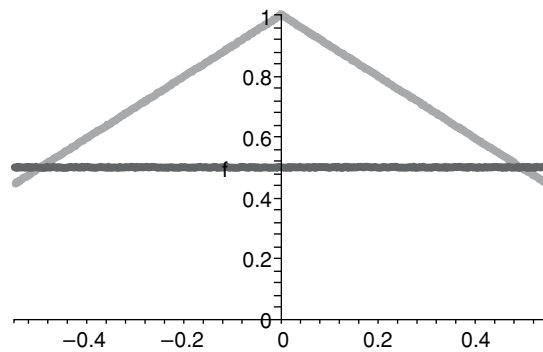
**Theorem 7.** *If  $Y$  is a discrete random variable of size  $n \geq 2$  with  $Y_{\min} = 0$  and  $Y_{\max} = 1$ , then  $\text{Dnt}(Y) \leq 1$ .*

*Proof.* Proof by induction. For  $n = 2$ , the theorem is correct by the result of the previous theorem. Assume that  $n > 2$  and that the theorem is true for  $2 \leq m < n$ . By definition, the dissent of  $Y$  is given by

$$\text{Dnt}(Y) = - \sum_{i=1}^n p_i \log_2(1 - |Y_i - \mu_Y|). \tag{A5}$$

Let  $Y_k$  be neither  $Y_{\max}$  nor  $Y_{\min}$ .

If  $Y_k$  is either zero or one, we can eliminate it, transfer its probability to either  $Y_{\max}$  or  $Y_{\min}$  without changing the value of  $\tilde{\mu}$ . Since this new distribution has size  $n - 1$ , by the induction hypothesis,  $\text{Dnt}(Y)$  is less than or equal to one.



**Figure A1.** When  $\mu = 0.5$ , then  $\mu^{\mu+\varepsilon} (1 - \mu)^{1-\mu-\varepsilon}$  is constant and below  $1 - |\varepsilon|$  over the interval  $-\mu < \varepsilon < 1 - \mu$ .

APPENDIX (Continued)

If  $p_k = 0$ , then, again  $Y_k$  can be eliminated from the summation, and this elimination does not affect  $\mu$ , so that by the induction hypothesis,  $\text{Dnt}(Y) \leq 1$ .

So we can now assume that  $Y_k$  is an interior point of  $Y$ —that is,  $0 < Y_k < 1$  with  $p_k > 0$ . We will now remove that term  $-p_k \log_2(1 - |Y_k - \mu|)$  from the summation in Equation 17, and then we will add, to the summation, two expressions, both of which are equal to zero, so that the value is not changed. These expressions are

$$p_k Y_k \log_2(1 - |1 - \mu|) - p_k Y_k \log_2(1 - |1 - \mu|) = 0 \tag{A6}$$

and

$$p_k(1 - Y_k) \log_2(1 - |0 - \mu|) - p_k(1 - Y_k) \log_2(1 - |0 - \mu|) = 0. \tag{A7}$$

If we collect terms in  $\text{Dnt}(Y)$ , then

$$\begin{aligned} \text{Dnt}(Y) &= -\sum_{i=1}^n p_i \log_2(1 - |Y_i - \mu|) \\ &= -p_k \log_2(1 - |Y_k - \mu|) + p_k Y_k \log_2(1 - |1 - \mu|) \\ &\quad + p_k(1 - Y_k) \log_2(1 - |0 - \mu|) - \sum_{i \neq k} p_i' \log_2(1 - |Y_i - \mu|), \end{aligned} \tag{A8}$$

where  $p_1' = p_1 + p_k(1 - Y_k)$ ,  $p_n' = p_n + p_k Y_k$ , and  $p_i' = p_i$  for  $1 < i \neq k < 2$ . By construction,

$$\begin{aligned} \sum_{i \neq k} p_i' &= [p_1 + p_k(1 - Y_k)] + \sum_{i=2}^{k-1} p_i + \sum_{i=k+1}^{n-1} p_i + (p_n + p_k Y_k) \\ &= \sum_{i=1}^n p \\ &= 1 \end{aligned} \tag{A9}$$

and

$$\sum_{i \neq k} p_i' Y_i = [p_1 + p_k(1 - Y_k)] Y_1 + \sum_{i=2}^{k-1} p_i Y_i + \sum_{i=k+1}^{n-1} p_i Y_i + (p_n + p_k Y_k) Y_n, \tag{A10}$$

but  $Y_1 = 0$  and  $Y_n = 1$ , so

$$\begin{aligned} \sum_{i \neq k} p_i' Y_i &= 0 + \sum_{i=2}^{k-1} p_i Y_i + \sum_{i=k+1}^{n-1} p_i Y_i + p_n + p_k Y_k \\ &= p_1 0 + \sum_{i=2}^{k-1} p_i Y_i + \sum_{i=k+1}^{n-1} p_i Y_i + p_n 1 + p_k Y_k \\ &= p_1 Y_1 + \sum_{i=2}^{k-1} p_i Y_i + \sum_{i=k+1}^{n-1} p_i Y_i + p_n Y_n + p_k Y_k \\ &= \sum_{i=1}^n p_i Y_i \\ &= \mu. \end{aligned} \tag{A11}$$

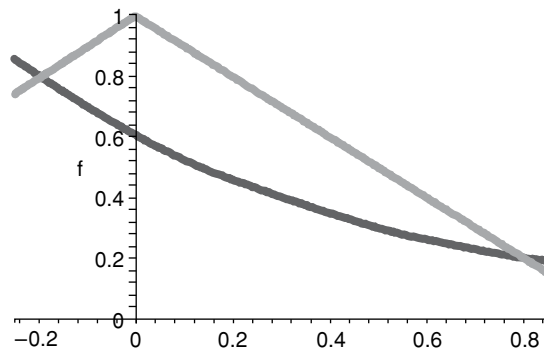


Figure A2. When  $\mu \neq 0.5$  then  $\mu^{\mu+\epsilon}(1 - \mu)^{1-\mu-\epsilon}$  is convex and below  $1 - |\epsilon|$  over the interval  $-\mu < \epsilon < 1 - \mu$ .

---

**APPENDIX (Continued)**


---

This says that the last term in Equation 17 is exactly in the form of a dissent of size  $n - 1$  and, by the induction hypothesis,

$$\begin{aligned} \text{Dnt}(Y) &\leq -p_k \log_2(1 - |Y_k - \mu|) \\ &\quad + p_k Y_k \log_2(1 - |1 - \mu|) \\ &\quad + p_k(1 - Y_k) \log_2(1 - |0 - \mu|) + 1. \end{aligned} \tag{A12}$$

This simplifies to

$$\text{Dnt}(Y) \leq -p_k \log_2(1 - |Y_k - \mu|) + p_k Y_k \log_2(\mu) + p_k(1 - Y_k) \log_2(1 - \mu) + 1. \tag{A13}$$

Let us examine the expression

$$T = -\log_2(1 - |Y_k - \mu|) + Y_k \log_2(\mu) + (1 - Y_k) \log_2(1 - \mu) \tag{A14}$$

and substitute  $\varepsilon = Y_k - \mu$ . Consequently, we have

$$\begin{aligned} T &= -\log_2(1 - |Y_k - \mu|) + Y_k \log_2(\mu) + p_k(1 - Y_k) \log_2(1 - \mu) \\ &= -\log_2(1 - |Y_k - \mu|) + Y_k \log_2(\mu) + (1 - Y_k) \log_2(1 - \mu) \\ &= -\log_2(1 - |\varepsilon|) + (\mu + \varepsilon) \log_2(\mu) + [1 - (\mu + \varepsilon)] \log_2(1 - \mu). \end{aligned} \tag{A15}$$

This expression is negative; by Corollary 1, multiplying by positive  $p_k$  does not change the sign of  $T$ , and we conclude that  $\text{Dnt}(Y) \leq p_k T + 1 \leq 1$ , so that the theorem is proved.

---

(Manuscript received January 3, 2006;  
revision accepted for publication June 26, 2006.)